DOCUMENT RESUME

ED 120 231 TM 005 190

AUTHOR Cross, Lawrence H.; Frary, Robert B.
TITLE A Study of Omitted Responses Under the Conventional Correction for Guessing.
PUB DATE Apr 76
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS Achievement Tests; *Guessing (Tests); Individual Characteristics; *Multiple Choice Tests; Personality Tests; *Response Style (Tests); Scores; *Scoring Formulas; Statistical Analysis; Undergraduate Students
IDENTIFIERS *Omitted Responses (Tests); Test Instructions

ABSTRACT

It has been demonstrated that corrected-for-guessing scores will be superior to number-right scores in providing estimates of examinee standing on the trait measured by a multiple-choice test, if it can be assumed that examinees can and will comply with the appropriate directions. The purpose of the present study was to test the validity of that assumption using data collected on 407 students enrolled in a freshman chemistry course at a large state university. The results showed that examinees who omitted items under formula scoring directions were omitting items for which, on the average, they were subsequently able to get significantly more correct than would be expected by chance when forced to guess. Moreover, there was a slight, but statistically significant relationship between success in guessing the correct answer to previously omitted items and two personality measures used in the study. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED120231

A STUDY OF OMITTED RESPONSES UNDER THE CONVENTIONAL CORRECTION FOR GUESSING

Lawrence H. Cross and Robert B. Frary
Virginia Polytechnic Institute
and State University

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Paper presented at the Annual Meeting of
the American Educational Research Association
San Francisco, California, April, 1976
(Session 25.07: Symposium, Division D)

TM005 190

ABSTRACT

A Study of Omitted Responses Under
the Conventional Correction for Guessing

Lawrence H. Cross and Robert B. Frary
Virginia Polytechnic Institute
and State University

It has been demonstrated that corrected-for-guessing scores will be superior to number-right scores in providing estimates of examinee standing on the trait measured by a multiple-choice test, if it can be assumed that examinees can and will comply with the appropriate directions. The purpose of the present study was to test the validity of that assumption using data collected on 407 students enrolled in a freshman chemistry course at a large state university. The results showed that examinees who omitted items under formula scoring directions were omitting items for which, on the average, they were subsequently able to get significantly more correct than would be expected by chance when forced to guess. Moreover, there was a slight, but statistically significant relationship between success in guessing the correct answer to previously omitted items and two personality measures used in the study.

A Study of Omitted Responses Under the Conventional Correction for Guessing

Lawrence H. Cross and Robert B. Frary*
Virginia Polytechnic Institute
and State University

Purpose

In a recent article, Lord (1975) demonstrated that to the extent that examinees can and will comply with the appropriate directions, then the conventional correction-for-guessing will provide a superior estimation of an examinee's standing on the trait being measured by a multiple-choice test than will number-right scoring. The explicit assumption made by Lord is that, when the appropriate instructions are provided, the only difference between an answer sheet obtained under formula scoring directions and the same answer sheet obtained under number-right scoring directions would be that omitted responses if any on the former answer sheet are replaced by random guesses on the latter.

While this assumption is far more tenable that the usual assumption associated with formula scoring, which is that examinees either know the answer and mark it or guess randomly among all item choices, it does not avoid the possibility that the admonition against wild guessing may cause examinees in general or examinees with certain personality characteristics to refrain from providing answers to questions which they have better than a chance probability of marking correctly, as suggested by several

---

*The authors wish to express their appreciation to Drs. John G. Dillard, Frederick K. Ross, Isabel Starling, Jimmy W. Viers, and their classes for their participation in the data collection phase of this study.

investigations (Votaw, 1936; Sheriffs and Boomer, 1954; Slakter, 1968). However, Lord has noted that certain technical difficulties evident in these studies do not permit an adequate evaluation of his assumption.

Another factor that may result in differential compliance with formula scoring directions is that some examinees may correctly judge that their expected score will not be adequate in a competitive examination and may therefore choose to answer all or nearly all questions in the hope that luck, in the short run, will be on their side. To the extent that these examinees are prevalent, the effect of Lord's assumption would be diminished.

Since there are at least two *a priori* reasons to suspect that certain students either cannot or will not comply with the directions appropriate for formula scoring, the purpose of this study was to test the validity of Lord's assumption in a typical test taking situation.

Instrumentation

The data for this study were obtained from three sections of a freshman chemistry course at a large state university. The three participating instructors developed a twenty item multiple-choice test to measure the mutually agreed upon course content. The items were selected from a computer based item pool and were screened by the investigators for item writing flaws. Each item contained four choices. In addition, four no-content (non-sense) items were written for which there was no correct or incorrect choice. One of the no-content items that was used is presented below:

Which of the following does not agree with Pauling's Rule for polyatomic gases?

a. Boyles Law
b. Charles Law
c. The combined gas law
d. Avogadro's Law

In as much as Pauling's Rule for polyatomic gases is fictitious, it may be assumed that if an examinee provided an answer to such a question, the answer would represent a sheer guess, the absence of information. Responses to such no-content items under formula scoring directions have been shown to be a reliable indicator of a personality trait referred to as utility-for-risk (Slakter 1967). These items were distributed among the twenty legitimate questions.

A second personality measure was obtained independently of the achievement test by having the students respond to the 39 items of the Minnesota Multiphasic Personality Inventory (MMPI) which constitute the "A-scale" as defined be Welsh (1956). Welsh suggests that the A-scale is a measure of general anxiety or general emotional upset. Scores are determined by the number of statements that a subject indicates are descriptive of himself with higher scores representing a higher degree of maladjustment. The items were administered approximately two weeks before the achievement test. Sheriffs and Boomer (1954) reported that examinees who scored high on the A-scale were penalized by formula scoring direction, and for this reason it was selected for use in this study.

Directions & Administration of Achievement Test

The directions for the achievement test were projected onto a large screen and were read aloud simultaneously to the examinees. The portion of the directions which explained the method of scoring and advised when to omit an item is presented below:

> Your score will be the number of items you mark correctly minus a fraction of the number you mark incorrectly. You should answer questions even when you are not sure your answers are correct. This is especially true if you can eliminate one or more choices as incorrect or have a hunch or feeling about which choice is correct. However, it is better to omit an item than to guess wildly among all of the choices given.

These directions were taken verbatim from Davis (1967) and are consistent with the recommendations made by Cureton (1966) and Lord (1975).

When the groups had had sufficient time to respond to the achievement test, they were directed to stop work and were asked to respond to three questions designed to assess their understanding of the directions and their basis for deciding when to omit an item.

Following this procedure, the examinees were provided with red pencils and were directed to answer all questions previously omitted. At this point, the examinees were told the purpose of the investigation and that their score for the test would in fact, be based on the number of correct choices marked, regardless of the color pencil used.

At the end of the ten week quarter, multiple-choice, final examination scores were obtained for the students in each of the three participating sections. Number-right scoring was used for this test.

Results

A preliminary comparison was made of the performance on the experimental test by examinees in the three class sections. Because of the similarity in performance in terms of the mean number-right scores, mean number of omitted items, and KR-20 estimates, the data from all three sections were combined for most of the analyses reported. There were 407 examinees for which data were available on both the MMPI and the experimental test. The descriptive statistics for all of the test scores, including the risk-taking measure, are presented in Table 1. From inspection of Table 1 it is obvious that it would be precarious at best to recommend one scoring method over another on the basis of differences in reliability and validity estimates.

The responses marked in red to items previously omitted under formula scoring directions were scored using the conventional correction-for-guessing formula.. (Resulting scores will be referred to hereafter as red-formula scores.) Were these responses to represent random guesses in the absence of information, the expected mean corrected score would be zero.

The observed mean red-formula score for those omitting items was .225 which is significantly different from zero at the .01 probability level ($t = 3.46$, $df = 240$). This result suggests that examinees of the type and in the setting described above do omit items for which they have a better than chance probability of answering correctly.

In order to avoid the assumption of a normal distribution for the red-formula scores, as required for the above test, a distribution-free test was made to determine whether the expected performance on omitted items differed significantly from chance expectation. For this analysis, answer sheets were separated according to the number of red marks and for each of the resulting nine groups, the probabilities associated with correctly answering exactly one, two, ... or nine items were computed. Both the expected and observed numbers of correct responses corresponding to each of the nine possible score values were summed over the nine groups. A chi-square, goodness-of-fit test revealed a significant difference between the observed and expected distributions ($\chi^2 = 15.4$, $df = 3$*).

Personality Measures and Red-Formula Scores

The correlation coefficient between scores on the A-scale of the MMPI and red-formula scores was -.03 when computed using all 407 examinees

*The observed and expected values associated with correctly answering more than three questions were very small and, therefore, were combined to yield a four level classification.

and was -.14 when computed using only the 241 subjects that had omitted at least one legitimate question on the achievement test. Inspection of the bivariate scattergram did not reveal a curvilinear relationship between these two variables. Only the latter coefficient of -.14 is significantly different from zero at the .05 probability level. This suggests that among those examinees who omitted legitimate items, those with lower scores on the A-scale (indicating better emotional adjustment) tended slightly to be more successful in choosing the correct choice to previously omitted items than those with higher scores.

The correlation coefficient between the measure of risk-taking and red-formula scores was -.12 for the entire group and -.06 for the restricted group of 241. The correlation of -.12 is statistically significant at the .05 probability level. When the examinees were grouped according to the number of no-content items to which an answer was provided, the 95 percent confidence intervals about the mean red-formula scores for the two lowest risk taking groups (those marking zero or one no-content items) did not contain the expected mean value of zero, whereas the 95 percent confidence interval did contain zero for the three groups of high risk takers. This result was observed for both the total sample and for the restricted group of 241. These results are presented in Table 2. The above analyses were interpreted to mean that there was a slight, but statistically significant, tendency for lower risk-taking students to be penalized by formula directions to a greater extent than high risk-taking examinees as measured by this index.*

*It should be noted that analyses of variance of the group means reported in Table 2 did not reveal significant differences. Thus, while the groups did not differ significantly from each other, there were significant differences when the observed group means were compared to expected mean values.

Compliance with Formula Scoring Directions

The response frequencies associated each of the choices to the three multiple-choice questions designed to assess understanding and compliance with formula scoring directions are presented in Table 3. Despite the fact that the directions were clearly projected on a large screen and simultaneously read aloud to a highly motivated college group, almost 27 percent of the examinees were not able to recall what the directions advised with respect to when a question should be omitted (choice c, question one).

A cross-tabulation of the responses to these three questions provided the basis for identifying two groups of examinees which are referred to hereafter as saints and sinners. Saints were identified as those subjects who 1.) understood the directions (choice c on question one), 2.) claimed to guess only if they could eliminate at least one choice (choice b on question two), and 3.) claimed to follow the directions as closely as possible (choice b on question three). The saints might be described as those in compliance with Lord's (1975) assumption. In contrast, sinners were identified as those examinees who also understood the directions but would provide an answer to a question even if it was a sheer guess among choices (choice c of question two) and knowingly guessed more frequently than suggested by the directions (choices b or c of question three). Using this classification there were 101 saints and 61 sinners.

Table 4 presents the means, standard errors, 95 percent confidence intervals and t values for these two groups on several variables. Inspection of Table 4 reveals that these two groups differed significantly in terms of their performance on the A-scale and the measure of risk-taking. It is interesting to note that those who complied with the directions were presumably

better adjusted psychologically than those who did not, as evident by the lower mean MMPI score for the saint groups. Whereas the mean number of omitted items was comparable for these two groups (1.89 vs. 1.73), Table 4 reveals that the 95 percent confidence interval associated with the mean red-formula scores for the saint group does not contain the expected mean of zero as in the case for the sinner group. Thus, even though these two groups did not differ significantly from each other in terms of red-formula scores, as evident by the t values, it appears that when the mean red-formula score for each group is compared to the hypothesized value of zero, the sinner group did slightly but significantly better than would be expected t chance. This finding may be interpreted as evidence that formula score directions are biased against those who understood and conscientiously attempted to comply with the directions.

Discussion

The data from this study suggest that examinees who omitted items under formula scoring directions were omitting items for which, on the average, they were subsequently able to get significantly more correct than would be expected by chance when forced to guess. If this result were to be replicated in other test situations, it would be difficult on an ethical basis to direct students to omit items even when the conventional correction-for-guessing formula is to be applied to the scores. Plausibly this outcome could have been caused by differing circumstances. Apparently 43 examinees interpreted the directions to conservatively (Table 3, choices 1a and 1b). Moreover, even though no one indicated certainty regarding the correct answer to all questions, 28 examinees claimed not to have guessed at all (Table 3, question two).

Another possibility is that examinees were not uniformly able to recognize circumstances under which they should have guessed even if they had understood and intended to comply with the directions. As Slakter (1968) suggests, it may be that examinees are simply unable to distinguish between complete, partial and no information. This conjecture is supported in the present study by the significant positive mean red-formula score for the 101 examinees identified as saints, who, by self report did understand and attempt to comply with the directions.

Personality differences may also identify groups more or less successful in marking previously omitted items. Success in guessing the correct answer to previously omitted legitimate items was found to be slightly related to both measures of personality used in this study. It may be argued that the differences observed in this study do not constitute substantial differences having practical implications, even though statistical significance was observed. However, on a longer test or with a test in a different subject matter, or with younger examinees, the number of omitted items may be considerably larger, and the statistically significant differences noted here may take on greater practical importance. Obviously, replications are needed before a general recommendation on this basis is warranted. None-the-less the data from this study did not support the assumption made by Lord even for those examinees who did understand and claimed to comply with the directions.

Quite aside from the outcomes just reviewed, it is discouraging to note that for a highly motivated college group, almost 27 percent of the examinees were not able to recall the response strategy suggested by the directions, and an even larger percentage was inconsistent in self-reported guessing behavior. Perhaps, as suggested by Lord, the difficulty can be corrected

by giving better test directions and if not, it may be time for the schools to teach children effective test taking behavior. Given the source from which the directions used in this study were taken (Davis, 1967), and the manner in which the directions were presented, it is doubtful, in the opinion of the writers, that better test directions could be formulated or that modifications would bring about the desired omitting behavior. Therefore, it appears that before the conventional correction-for-guessing is routinely adopted, schools will have to meet Lord's challenge to teach children more effective test taking behavior.

REFERENCES

Cureton, E.E. The correction for guessing. *The Journal of Experimental Education*, 1966, *4*, 44-47.

Davis, F.B. A note on the correction for chance success. *The Journal of Experimental Education*, 1967, *3*, 43-47.

Lord, F.M. Formula scoring and number-right scoring. *Journal of Educational Measurement*, 1975, *12*, 7-11.

Sheriffs, A.C., & Boomer, D.S. Who is penalized by the penalty for guessing? *Journal of Educational Psychology*, 1954, *45*, 81-90.

Slakter, M.J. The penalty for not guessing. *Journal of Educational Measurement*, 1968, *5*, 141-144.

Votaw, D.F. The effect of do-not-guess directions upon the validity of true-false or multiple choice tests. *Journal of Educational Psychology*, 1936, *27*, 698-703.

TABLE 1
Test Statistics

Test	n	$\bar{x}$	S.D.	KR-20
MMPI A-Scale	407	12.00	6.91	.87
Achievement Test				
a. Number right scores	407	10.12	3.44	.655
b. Formula scores	407	9.59	6.74	.664
Risk Taking Measure	407	2.84	1.27	.649
Final Examination*				
Section 1 (38 items)	160	25.6	5.59	.798
Section 2 (40 items)	139	26.6	5.68	.791
Section 3 (40 items)	150	23.7	7.22	.869

Validity Coefficients

Final Examination	Achievement Test: Number Right Scores	Achievement Test: Formula Scores
Section 1 (n = 138)	.468	.464
Section 2 (n = 101)	.604	.622
Section 3 (n = 131)	.583	.590

*A different set of questions was used in each section; and within each section, parallel forms were generated by scrambling the items. These statistics, therefore, represent average values over two or more parallel forms.

TABLE 2

Red-formula Score Performance for Examinees Classified According to Level of Risk-taking.

All Examinees (n = 407)

Number of No-content Items Marked	n	$\bar{x}$	95% C.I.
0	27	.37	.037 to .705
1	45	.39	.012 to .759
2	66	.01	-.181 to .212
3	97	.16	-.019 to .336
4	172	.06	-.013 to .134

Only Those Examinees Having Omitted Legitimate Items Under Formula Directions (n = 241)

Number of No-Content Items Marked	n	$\bar{x}$	95% C.I.
0	25	.40	.041 to .760
1	40	.43	.014 to .853
2	51	.02	-.236 to .276
3	78	.20	-.024 to .418
4	47	.22	-.050 to .491

TABLE 3

Response Frequencies to Three Questions Designed to Assess Understanding of and Compliance with Formula-Scoring Directions.

1. According to the directions at the start of the test, you were advised:

29	a.	to omit answering a question if you were not certain of the answer.
14	b.	to omit answering a question if you could eliminate only one of two choices as incorrect.
298	c.	to omit answering a question if you had to guess among all choices.
66	d.	not to omit answering any questions under any condition.

2. Quite apart from what the directions suggested, what did you usually do when you were not certain of the answer to a question?

28	a.	I refrained from guessing.
238	b.	I guessed only if I had a hunch as to the correct answer or could eliminate at least one choice as incorrect.
141	c.	I provided an answer even if my answer was a sheer guess among all the choices.
0	d.	(I was certain of the correct answer to all questions.)

3. While taking this test, it was my personal decision to:

171	a.	follow the directions as I understood them as closely as possible.
62	b.	guess more frequently than suggested by the directions as I understood them hoping that luck would be in my favor, although I believe I would have made a satisfactory grade without additional guessing.
126	c.	guess more frequently than suggested by the directions as I understood them because otherwise I doubted my ability to get a satisfactory grade on this test.
48	d.	guess less frequently than suggested by the directions as I understood them in order to avoid losing score points for wrong answers.

TABLE 4
Test Performance of Saints vs. Sinners
(n = 101 and 61 respectively)

	$\bar{x}$	$s/\sqrt{n}$	95% C.I.	t value
MMPI				
saints	9.79	.591	8.61 to 10.97	3.69*
sinners	13.00	.886	11.23 to 14.77	
Risk-taking				
saints	2.37	.118	2.13 to 2.60	3.99*
sinners	3.11	.135	3.04 to 3.58	
Number Omitted				
saints	1.89	.191	1.51 to 2.27	.08
sinners	1.73	.260	1.22 to 2.26	
Red-formula Scores				
saints	.188	.083	.024 to .353	.69
sinners	.099	.083	-.067 to .264	

*$p<.01$